DOCUMENT RESUME

ED 177 207 TM 009 709

AUTHOR Strand, Theresa
TITLE Who Said What About Title I Evaluation at the AERA/NCME Annual Meeting, Toronto, March, 1978: Summaries of Selected Papers.
PUB DATE Oct 78
NOTE 14p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Gains; Compensatory Education Programs; Educational Testing; Elementary Secondary Education; *Evaluation Methods; *Models; *Program Evaluation; Research Design; *Research Problems; Testing Problems
IDENTIFIERS Elementary Secondary Education Act Title I; *RMC Models; *Title I Evaluation and Reporting System

ABSTRACT

Twenty-six papers concerning Elementary and Secondary Education Act Title I program evaluation--presented at the 1978 American Educational Research Association/National Council on Measurement in Education meeting--are summarized. Speakers from the federal government are: C.T. Cross (see ED 156 695); B.L. Fishbein; and D. Ozenne. State department of education speakers are D.R. Kearns of Arizona and R.O. Maye of Illinois. Speakers from local school districts are J.A. Arter and G. Estes (ED 170 332); C.R. Crowder and E.J. Gallas (ED 155 212); S. Powers and E.J. Gallas (ED 155 211); H.B. Slaughter and E.J. Gallas (ED 155 213); and S.H. Davidoff and C.R. Iwamoto. University representatives include F.X. Archambault and R.G. St. Pierre (Ed 152 953); R.M. Jaeger (ED 164 550); R.L. Linn (ED 156 696); G.K. Mandeville (ED 159 224); J. Long et. al. (ED 161 914); and D.E. Wiley. Speakers from the RMC Research Corporation are L.A. Dougharty and L.S. Kilibancff; and G.K. Tallmadge and D.P. Horst (ED 155 204). Representatives of regional technical assistance centers are B.L. Anderson; G. Echternacht; G.D. Estes and J.I. Anderson (ED 164 601); J. Haenn and D.C. Proctor (ED 171 728); J.B. Hansen; S.B. Hiscox and T.R. Owen (ED 166 258); A.J. Stenner et. al. (ED 169 072); and K.O. Yap (ED 167 622). (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED177207

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

WHO SAID WHAT ABOUT TITLE I EVALUATION AT THE AERA/NCME ANNUAL MEETING, TORONTO, MARCH 1978: SUMMARIES OF SELECTED PAPERS

Theresa Straud,
Educational Testing Service, Midwestern Regional Office
Region V Technical Assistance Center for Title I Evaluation

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Theresa Straud
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

AT THE FEDERAL LEVEL:

*Cross, Christopher T. (Minority Staff Director, Committee on Education and Labor, U. S. House of Representatives), "Title I Evaluation-A Case Study in Congressional Frustration."

Traced the history of Section 151, which required OE to develop evaluation models, and discussed the potential usefulness of the models developed. Highlights of the presentation follow.

There has been considerable discussion with respect to the feasibility and technical problems created by the RMC Title I evaluation models. Although issues of general concern have been raised, Congress is generally not concerned with technicalities in the development of evaluations. Congress wants information upon which to base policy and political judgments about programs. The primary concern of Congress is that information produced can withstand a professional critique and that data will produce findings which are at least somewhat conclusive.

Although there is some truth to the charge that decisions of Congress are essentially political and rarely based on data or policy consideration, data and good solid information can have considerable impact on what Congress does. For example, a highly critical General Accounting Office study resulted in Congress restructuring many portions of the Vocational Education Ammendments of 1976. Also, as a result of the NIE compensatory education study, there has been a proposed revision and reorganization of the Title I statute.

Information obtained from use of the RMC Models is certain to have an impact on the actions of future Congresses with respect to the design of Title I. More important is the impact that evaluation findings should have on the operation of Title I programs at the local and state levels.

In the final analysis, if the models are even 50 percent effective in bringing to the attention of school officials and parents the successes and failures in Title I programs, the investment in the models and implementation of Section 151 will have been repaid many times over.

*Fishbein, Ronald L .(U. S. Office of Education, Office of Planning, Budgeting and Evaluation), "The Use of Non-Normed Tests in the ESEA Title I Evaluation and Reporting System, Some Technical and Policy Issues."

Identified and discussed various technical and policy issues

TM009 709

Identified and discussed various technical and policy issues related to the use of non-normed tests in the ESEA Title I Evaluation and Reporting System.

Technical issues discussed were: (1) the assumption, necessary to implement Models B2 and C2, that the ratio between the sample and population standard deviations on the nationally normed test and the test without national norms is the same (assuming the same population and sample for each test); (2) the linear transformation used in Models B2 and C2 to convert the local treatment effect to NCEs, along with possible alternative procedures for using non-normed tests in the Title I system; and (3) floor and ceiling effects in relation to Model A2 evaluations.

Policy issues discussed included the scope of the curriculum possible to test when using Model A2, B2, or C2 evaluation, and additional procedures that might be undertaken. It was concluded, that in the vast majority of cases, only a limited set of program objectives can be evaluated with a standard Model A2, B2 or C2 implementation. It was recommended that those employing a non-normed version of the Title I evaluation system consider incorporating a normative assessment component.

*Ozenne, Dan (U. S. Office of Education), "The Effect of Vertical Scaling Imprecision in the Estimation of Title I Project Gains."

Discussed the occurrence of vertical scaling errors, suggested reasons for their occurrence, and recommended procedures to minimize their effects. Highlights of the presentation follow.

Considerable research evidence suggests that expanded score scales may contain considerable error. Depending on particular tests and level used, vertical equating problems may enhance or depress any actual treatment effects. Model A results may be suspect when: (1) a below grade level test is used for pretest and an on-grade level test is used for posttest, or (2) in year-to-year evaluation designs based on different test levels, each appropriate to the grade level tested.

Some possible causes of scaling error are: (1) test levels may differ in instructional content, and (2) test levels may differ in the norming and equating processes, for example, in size and composition of norming sample and in the amount of smoothing or curve fitting done at distribution extremes where Title I pupils are located. Several recommendations include: (1) avoid changing test levels whenever possible by selecting for content (appropriate test spans the required time frame and has sufficient range to eliminate floor and ceiling effects; longer tests have a greater measurement range), and (2) determine if there is a test level artifact present in program results.

AT THE STATE DEPARTMENT LEVEL:

*Kearns, Donald R. (Arizona Department of Education), "ESEA Title I Evaluation Model Equivalence and Comparability With Real Data."

Compared Title I evaluation models and identified their relative strenghts and weaknesses. Reported on a study conducted to assess the validity of equivalence of project results regardless of evaluation models used. A high degree of variability in NCE gains across models was found.

*Maye, Rose Owens (Program Evaluation, Illinois Office of Education) "The Pecularities of Title I Evaluation."

Identified and discussed "peculiarities" of Title I evaluation (1) the heavy reliance on test scores to evaluate programs does not provide the overall prespective needed to adequately and fully evaluate a program; (2) the new Title I evaluation system makes reporting to parents increassingly difficult because of NCE scores and lack of information on instructional and supportive services; (3) program evaluation should be used to examine affective and supportive services to determine how they affect the Title I students' academic performance; and (4) the new Title I evaluation system covers programs that start in grade two whereas many Title I programs focus on per-kindergarten to grade 3. Solution suggested was to include descriptive, judgmental, and objective test data as equally weighted components of Title I evaluation.

AT THE SCHOOL DISTRICT LEVEL:

*Arter, Judith A. (Phoenix Union High School System), and Estes, Gary D. (Northwest Regional Educational Laboratory), "A Model for Developing Local Norms with a Standardized Achievement Measure for Use with Local Program Evaluation: Procedures & Effects."

The Phoenix Union High School System instituted two procedural changes: (1) October-to-October testing to ensure that judgments concerning program efficacy are not based on fall-to-spring gains which may disappear over the summer, and (2) development of local norms by grade level and ethic group. Local group norms were found to provide a better comparison group than national norms. Also, ethic group norms were found to be useful for program evaluation and for determining whether the test had ethnic bias. In addition, Arter and Estes examined two other issues: (1) validity of CTBS for local use, and (2) problems of progressively better ability students in the norming sample. The CTBS was found to be a valid measure of achievement, inasmuch as the test correlated highly with the number of courses taken, with district-made tests, and student ability level.

*Crowder, Christopher R. (Research and Evaluation Deparment, Tucson Unified School District), and Gallas, Edwin J. (Data Analysis) "Relation of Out-of-Level Testing to Ceiling and Floor Effects On Third and Fifth Grade Students."

Reported on a study which investigated the relation of out-of-level testing to ceiling and floor effects on third and fifth grade students. Concluded that the comparability of standard scale scores between the on-level and out-of-level tests appeared to relate more to the test being used, its raw score-scale score transformation, and the relative difficulty of the test, than to the previous floor and ceiling effects shown by students tested on-level.

*Davidoff, Stephen H. (Assistant Director, School District of Philadelphia), and Iwamoto, Carrolyn K. (Research Assistant, School District of Philadelphia), "New Trends in Title I Evaluation: Will Aggregation Cause Aggravation? (A Tale of Three Models)."

Reported on Philadelphia's plans for Title I evaluation. The school district plans to bypass the building level in the collection of information and aggregate data centrally. Reasons are: (1) computer files have funding information, (2) training teachers to do evaluations would be expensive, (3) individual pupil test scores can be obtained centrally from city-wide test files and a tracking system is being developed, and (4) the district employs Model A as the result of a comparative study they conducted of the three models using the same group of students from 12 Title I elementary schools. The best possible results were achieved with Model A, then Model B. Model C showed no overall impact.

*Powers, Stephen (Department of Research and Evaluation, Tuson Unified School District), and Gallas, Edwin J. (Data Analysis, "Implications of Out-of-Level Testing for ESEA Title I Students."

Reported on a study that was conducted whose purpose was to: (1) compare achievement of students tested with on-level and out-of-level reading comprehension subtests of the CTBS/S, (2) determine if on-level or out-of-level subtests were suitable by investigatng the chance level, floor and ceiling effects, test suitability and a reliability index, (3) investigate any significant differences between mean expanded scores, and (4) study any linear or curvilinear trends in the data.

Results showed an increase in measurement precision obtained through testing students at their instructional (functional) level. Also, results showed that obtained expanded standard score means from on-level and out-of level tests were reasonably equivalent within sampling error. Concluded that if a small bias is introduced into the data by testing out-of-level, it is preferable to accept this small bias if it is therefore possible to increase substantially the precision of measurement. The possibility of curvilinear trends was investigated but no such trends were found.

*Slaughter, Helen B. (Department of Research and Evaluation Tucson Unified School District) and Gallas, Edwin, J. (Department of Research and Evaluation, Tucson Unified School District), "Will Out-of-Level Norm-Referenced Testing Improve the Selection of Program Participants and the Diagnosis of Reading Comprehension in ESEA Title I Programs?"

Reported on a study whose major purpose was to test the assumption that groups of students in low achieving schools would obtain the same interpretive score (for example, percentile rank and NCE) when tested on reading tests one or more years below their grade placement level as they would have obtained on on-grade level tests.

Study utilized CAT, C to test grade 6 pupils in Title I schools. Findings indicated that students obtained lower derived scores when tested below level than when tested at grade placement level. It was recommended that out-of-level empirical norms be developed for out-of-level testing by testing students at multiple test levels in norming procedure.

AT THE UNIVERSITY LEVEL:

*Archambault, Francis X. (University of Connecticut), and St. Pierre, Robert G. (Abt Assoiates Inc.), "A Description of Instructional Experiences Received By Title I and Non-Title I Students."

Developed an instrument, "Classroom Activities Log (CAL)." designed to obtain instuctional experience data. Reported on a study which used the instrument to provide information on instructional experiences of Title I and Non-Title I pupils.

*Jaeger, Richard M. (University of North Carolina at Greensboro), "The Effect of Test Selection on Title I Project Impact."

Maintained that Congress is unlikely to respond to nationwide

data on the impact of Title I and that the present USOE-RMC plan for aggregating achievement gains across Title I projects is technically unsound. Jaeger suggested an alternative plan for aggregating findings on the attributes of project impact. For example, attributes would be well-defined global categories with descriptions like "highly effective," "effective," "ineffective," and "aversive." Each of the categories would be defined, for example, "highly effective" might be a project in which 75 percent of the participants achieved project goals. Nationwide statistics could state: "Of all Title I reading projects for elementary school children, 62 percent were judged to be highly effective." The proposed system would place primary emphasis on local evaluation, requiring evaluation for informed local project decisions, and imposing evaluation standards tailored to the design, purpose, and context of individual projects.

*Linn, Robert L. (University of Illinois at Urbana-Champaign), "The Validity of Inferences Based on the Proposed Title I Evaluation Models."

Focus of discussion was on issue of validity, particularly internal validity of RMC models, with emphasis on model A. Questions considered regarding Model A were: (1) is there an adequate basis for using a constant percentile or the no-treatment expection -- evidence was presented that the answer is no; and (2) is it reasonable to use norms from one test to establish the expected no-treatment performance level for another test (in reference to use of normed and non-normed test) -- evidence was presented that the answer is again no.

Other highlights of Linn's presentation follow: (1) Model B ranks highest of the three models in internal validity but it has operational problems (cost feasibility) and ethical objections, (2) Model B can only be approximated and statistical adjustments rest on strong assumptions, and (3) Model C is potentially useful in specialized circumstances but its validity depends on strong assumptions especially that of linearity.

*Mandeville, Garrett K. (University of South Carolina), "An Evaluation of Title I Model C1, The Special Regression Model."

Presented a systematic study of Model C1 characteristics and

compared Model C1 to the norm-referenced model. Mandeville found that the results did not support the expectation of unbiasedness of SRM estimates. However, he found bias not large, particularly when compared with bias of NRM estimates.

*Long, John (University of Rhode Island), Horwitz, Stephen (University of Rhode Island); DeVito, Pasquale (Rhode Island State Department of Education), "An Empirical Investigation of the ESEA Title I Evaluation Systems' Proposed Variance Estimation Procedures For Use With Criterion Referenced Tests."

Raised issues concerning the assumption that the population standard deviation of a non-normed test can be accurately estimated since "the ratio of the local to the national standard deviation will be the same for the normed test as for the non-normed test". Conducted a study of 318 Title I student and found differences between estimated and actual national standard deviations that could understate or overstate effects of a Title I program. Long suggested that a study be conducted to determine: (1) the degree to which errors are systematic or random (2) the magnitude of errors which might be expected, and (3) a system for establishing confidence bands around the estimated standard deviations.

*Wiley, David E. (University of Chicago and ML-GROUP for Policy Studies in Education, CEMREL, Inc.), "Evaluation by Aggregation: Social and Methodological Biases."

Suggested that technical issues of Title I evaluation may be secondary to political issues: If Title I is intended to increase the reading competencies of educationally deprived pupils, a valid assessment of the corresponding level of accomplishment, if feasible, is a valuable political tool.

Identified and discused some relevant questions: (1) what are the "basic" competencies to be assessed, (2) are contents of tests sufficiently homogeneous to allow their conversion to a meaningful common scale, (3) which of those potentially possible score conversions are actually feasible with existing data and procedures, and (4) have the evaluation data to be transformed been collected using procedures and designs which permit valid references about the effects of the Title I program.

Analyzed information needs pertaining to title I evaluation and identified three important kinds of evaluation outcomes: (1) assessments of the degree of effectiveness of local programs, (2) assessments of the overall impact of Title I as a national program, and (3) relationships between program characteristics, components or features, and effectiveness. Willy compared these factors to be critical for selecting projects for possible packaging and replication, and for forming legislative mandates, agency regulations, and guidelines for program types and features leading to successful outcomes.

Willy also suggested that greater emphasis be placed on improving the reporting system by which program characteristics and outcomes are accumulated and shifting analyses of them to OE as part of the "aggregation" effort.

AT RMC:

*Dougharty, Laurence A. (RMC Research Corporation), and Kilibanoff, Leonard S. (RMC Research Corporation), "A Policy Analysis of the 'Critical Mass' Concept for the Distribution of Compensatory Education Funds."

Reported on an evaluation of the Critical Mass Policy for allocating compensatory education funds in California. Highlights of the presentation follow.

Under the Critical Mass Policy, school districts must spend a minimum of $350 per participating student. An educational production analysis indicated that there was no relationship between the level of concentration of compensatory funds and increase in achievement test scores. An analysis of successful and unsuccessful schools yielded the same conclusion. It can be implied that the Critical Mass Policy has not been successful and that funds should be distributed by criteria other than expected gain in achievement scores. Critical Mass Policy is in need of revision to provide for more local control.

* Tallmadge, G. Kasten (RMC Research Corporation), and Horst, Donald P. (RMC Reserch Corporation), "The Use of Different Achievement Tests in the ESEA Title I Evaluation System."

While not denying the problems inherent in the use of many

different tests, Tallmadge maintained that allowing LEAs to select their own tests substantially increases the precision of the system as compared to the alternative of a single, system-wide test battery. The important thing, he maintained, is to match content of the test to the Title I program. He defended the new evaluation system by using an analogy to atheletes: a ski jumper how is three stand ard deviations above the mean of competitive ski jumpers can be considered a better athlete than a high jumper who is only 2.9 standard deviations.

AT THE TAC CENTERS:

*Anderson, Beverly L. (Technical Assistance Center, Northwest Regional Educational Laboratory), "A Look At Empirical Norms."

Discussed various sources of error which could affect Model A1 evaluation: (1) lack of test reliability, (3) errors introduced through additional score conversions required for out-of-level testing, (4) errors that may occur when scores are converted from one scale to another, (5) improper test administration, and (6) errors due to norms. Anderson also identified various types of norms error: (1) norming sample not accurately representing population in norm group, (3) different norming samples used for pre- and posttesting, and (4) use of extrapolated or interpolated norms. Anderson emphasized that the relative importance of these various sources of errors remains unanswered.

Other highlights of the presentation follow. Once content validity has been established, one must determine the extent to which Title I students resemble norms sample on the following: (1) size of school, school districts, cities, (2) geographical region, (4) race or cultural characheristics, and (5) socioeconomic status. It is also necessary to determine factors relating to achievement of norm groups, such as: (1) use of empirical or interpolated data, (2) use of random sampling procedure to select norms, (3) if sampling non-random, evidence to indicate that sample represent the population it claims to represent, (4) testing of norms group within a short span of time, for example, two-form week period.

Suggested studies include comparing results between interpolated norms and empirical norms when: (1) using the same students at two times during the year, and (2) when using different samples of students for those same two times during the year.

*Echternacht, Gary (Educational Testing Service), "The Use of Different Models in the ESEA Title I Evaluation System".

Maintained that no theoretical or empirical evidence has been advanced to support USOE's claims that all models produce unbiased and comparable results. Presented findings from a computer simulation study comparing results with Model A with those of Model C. Data was created through Monte Carlo techniques to simulate test and learning behavior in order to control variables such as time between testing, measurement process, growth patterns, and achievement gains. Highlights of the presentation follow.

Results of the simulation suggested that comparable results will not be produced by the system because: (1) except for a situation where pretest is difficult and posttest easy, Model A is biased, (2) Model A bias results from norms group sampling, moderately difficult pre- and posttests, and continuous measurement, (3) Model A bias is large in relation to standard errors for both Model A and C, (4) mean square errors for Model C are smaller than those for Model A, indicating that they give a better estimate of treatment effect, and (5) Model A does not correct for any regression to the mean.

Since most school districts are using Model A, the resulting aggregated NCE gains most likely will be overestimates of true gain. Also, districts using Model C will be penalized if funding decisions are based solely on NCE gain.

Recommended that USOE give up aggregating NCE gains and (1) set up specific targets, perhaps an NCE gain of three points if Model A were used and perhaps a positive NCE gain if either Norm C or a modified Model A were used, and (2) gradually phase out Model A and use only Models B and C.

Districts with district-wide testing program have no problem adopting Model C; districts without, are forced to use Model A. Suggested that these districts could use Model C if pretest did not have to be at the beginning of project, and if selection data could be more than one year old.

*Estes, Gary D. (Northwest Regional Educational Laboratory) and Anderson, Judith I. (The BDM Corporation), "Observed Treatment Effects with Special Regression Evaluation Models in Groups with No Treatment."

Reported on an empirical study to obtain treatment effect

estimates using the Special Regression Model for groups in which there was no treatment. Results for the Title I Special Regression Model were calculated for about 730 ninth graders enrolled in general mathematics classes in a large city school system. Conclusions were: (1) even minor ceiling or floor effects resulted in nonhomogeneous regression lines for treatment and comparison groups, (2) tests of statistical significance should precede conclusions of educational significance at the local school or district level, and (3) guidelines to assist with interpreting evaluation results are needed.

*Haenn, Jospeh (NTS Research Corporation), and Proctor, Diane C. (NTS Research Corporation), " A Practioner's Guide to Out-of-Level Testing."

Reviewed literature and events which have had impact on concept of out-of-level testing and explored reasons for the following Testing out-of-level, procedures to be followed, and required steps for converting out-of-level scores to appropriate in-level norms. Appendix A of paper provides chance level scores and their percentiles and grade equival ents for the commonly used achievement tests, Appendix C presents two alternative work sheets for use in score conversion procedures.

*Hansen, Joe B. (Northwest Regional Educational Laboratory), "Measurement Problems in Title I Evaluation."

Discussed four key issues relating to Model A1: (1) equipercen tile assumption, (2) regression phenomenon, (3) out-of-level testing, and (4) use of various empirical and non-empirical norms studies on equipercentile assumption. Hansen concluded that the equipercentile assumption appeared to be adequate in predicting spring scores in fall-to-spring evaluations but there was a tendency for expected posttest scores to be underestimated for those with extremely high pretest scores; and that the validity of the equipercentile assumption depended on the repres entativness of norms with respect to Title I group. Hansen identified a number of student selection alternatives to avoid systematic bias introduced by selection/pretest regression effect.

Hansen discussed various problems relating to out-of-level testing. He maintained that empirically derived norms vary with respect to national representativeness and suggested that extrapolated norms from a truly representative sample might be better than empirical norms from a less representative sample

Other problems he identified relating to norms include the following: (1) norms groups might not provide an adequate norming base for Title I pupils who tend to have a disproportionate representation of ethnic minority and low SES pupils, and (2) inclusion of Title I pupils in the norming sample produces a built-in treatment effect in the norms.

*Hiscox, Suzanne B. (Northwest Regional Educational Laboratory) and Owen, Thomas R. (Educational Consultant, Portland, Oregon), "Behind the Basic Assumption of Model A."

Found data from previously conducted studies to be inclusive in terms of the equipercentile assumption. Hiscox and Owen identified various factors affecting the accuracy of the assumption. These are (1) degree of similarity of norm group to Title I students, (2) concomitant with dropout of low scoring students in high school, the score needed to achieve a given percentile level from on testing time to another will increase, (3) the procedures used by publishers to establish norms (use of cross-sectional rather than longitudinal groups of students in developing test norms) present potential problems in articulation between levels of a test battery and smoothing of growth curves, (4) out-of-level testing, (5) floor and ceiling effects, (6) test selection, and (7) two different types of regression.

The authors reported on a study they conducted which was designed to examine the equipercentile assumption using longitudinal data from Portland Public Schools. Findings were inclusive because it was difficult to determine wherther variability was due to real differences in student achievement, to changes in test levels administered over four years, or to incomplete test data.

*Stenner, A. Jackson (NTS Research Corporation), Bland, June D. (Washington D. C. Public Schools), Hunter Earl L. (Washington D. C. Public Schools), Cooper, Mildred L. (Washington D. C. Public Schools), "The Standardized Growth Expectation: Implications for Educational Evaluation."

Discussed assumptions underlying use of norm-referenced tests: (1) treatment effects expressed in a standard score metric permits aggregation of effects across grades, (2) commonly used standardized tests are sufficiently comparable to permit aggregation of results across tests, and (3) summer loss in achievement skills and knowledge. Presented a number of hypotheses related to these assumption and suggested needed research.

*Yap, Kim Onn (Northwest Regional Educational Laboratory), "Can Selection Tests be Used as Pretests?."

Reported on a study conducted: (1) to estimate the severity of regression bias when a selection test is also used as a pretest, and (2) to assess the performance of existing proced ures for reducing or eliminating bias. Yap found that, overall, the regression bias does exist and that size of bias was a direct function of data reliability and the "extremeness" of the subgroup. Yap presented an adjustment method which predicts pretest scores on the basis of selection and the correlation coefficient between selection and pretest, with the correlation coefficent being replacable by such indices of reliability as the internal consistency coefficient.